

February 23, 2024

E-Filed

The Honorable Donna M. Ryu
United States District Court for the Northern District of California
Oakland Courthouse, Courtroom 4 – 3rd Floor
1301 Clay Street
Oakland, California 94612

Re: *Doe I, et al. v. GitHub, Inc., et al.*, Lead Case No. 4:22-cv-06823-JST-DMR

Dear Judge Ryu:

Pursuant to ECF. No. 175, Paragraph 14 of this Court’s Standing order, and the Court’s prior order regarding this discovery dispute, ECF 212, the parties submit this discovery joint letter brief. Defendants Microsoft and GitHub request that the Court issue a protective order delaying discovery into Microsoft’s 2018 acquisition of GitHub (the “Acquisition”) until such time, if ever, Plaintiffs make a *prima facie* showing they are likely to certify a damages class. Counsel met and conferred in person on February 13, 2024, but were unable to resolve the dispute.

I. Current Case Deadlines

(1) Fact Discovery, September 27, 2024; (2) Expert Discovery, February 21, 2025; (3) Dispositive Motion Hearing (not scheduled); (4) Class Certification Motion, March 27, 2025; (5) Pretrial Conference and Trial Dates (not scheduled).

II. Defendants’ Position

Plaintiffs served two interrogatories, nine RFPs, six subpoenas, and three letter briefs seeking a broad array of discovery regarding the Acquisition, including identification of all people involved and production of all documents, communications, and diligence materials. Ex 1.

The only claims presently in this case after the Court’s orders on motions to dismiss are two contract claims alleging (1) breach of open source software (“OSS”) licenses requiring identification of authors when code is copied and distributed, and (2) breach by GitHub of its terms of service promising not to sell personal data. In addition to limiting the substantive claims, the Court has also limited standing, holding that only three of the five named plaintiffs have demonstrated sufficient actual injury to sustain a damages claim.

The elements of a claim for breach of contract are formation, plaintiff’s performance, defendant’s breach, fact of injury, and damages. *See* CACI 300, 303; BAJI 10.85. The measure of damages in a case for breach of contract is benefit of the bargain, *i.e.*, expectation damages. *Applied Equipment Corp. v. Litton Saudi Arabia Ltd.*, 7 Cal.4th 503, 515-18 (1994); *see* Cal. Civ. Code § 3300. Unjust enrichment is not an available remedy for breach of contract, where damages are available as an express remedy of law. “No action for unjust enrichment lies where a contract governs the parties’ relationship to each other.” *Talavera v. Global Payments, Inc.*, No. 21-CV-1585 TWR, 2023 WL 3080701, at *21 (S.D. Cal. Apr. 25, 2023).

February 23, 2024

Page 2

Motive is irrelevant to a breach of contract claim. *Applied Equip.*, 7 Cal.4th at 516 (motives for breach are “immaterial”); *Rich v. Shrader*, No. 09–CV–0652–MMA (WMC), 2010 WL 3717373, *4 (S.D. Cal. Sept. 17, 2010) (“Motive is not an element of a breach of contract action”); *Davis v. Leal*, 43 F.Supp.2d 1102, 1112 (E.D. Cal. 1999) (“For the contractual remedies, Leal's motive or other state of mind in breaching the agreement is not pertinent to the discussion”).

Count 2 is a claim for breach of OSS licenses by failing to attribute code outputs to the Plaintiffs. Count 3 is a claim for breach of GitHub’s terms of service claiming that GitHub “sold” Plaintiffs’ personal data in the form of the public code repositories. All of the code at issue is by definition licensed without charge. As it is free code (free as in *gratis*, not free as in freely available), it makes no sense to assert that Microsoft and GitHub selling a code-completion service that hypothetically spits out snippets of Plaintiffs code could harm the value *of that code*. Anyone can go and get it for free at any time. Rather, the benefit of the bargain was that Plaintiffs would get credit as authors. The injury is the lack of additional benefit to their reputation. Thus, even amongst the named Plaintiffs (let alone a purported class of OSS licensees), each reputation must be examined and the quantum of lost benefit determined. There is no possibility of a uniform injury or theory of damages. Each Plaintiff will have to separately prove not only that there were matching outputs, but also explain how such outputs caused each one of them harm.

Microsoft’s acquisition of GitHub is irrelevant to such theories of breach. The Acquisition is between Microsoft and GitHub; it has nothing to do with the terms of contracts agreed between GitHub and its customers, or between Microsoft and OSS licensors, whether any GitHub Copilot output matches Plaintiffs’ code, whether any such output has in fact caused reputational harm, or how to ascertain the amount of injury, if any, caused to each of the Plaintiffs. The Acquisition simply has nothing to do with any of these elements of the claims.

Plaintiffs have changed their relevance arguments at least five times during the process of this motion. One theory is that the acquisition value of GitHub, Inc. is somehow a proxy for all of the value of the lost attribution. This is preposterous. The value of all of GitHub, Inc. bears no relationship to any individual Plaintiffs’ reputation, all of the class of Plaintiffs’ reputations, or even the characteristics, demand for, or qualitative value of the purported class’s code. If it is supposed to be a proxy for the value to Defendants of the models, that is an unjust enrichment theory that is not available on a breach of contract claim.

Plaintiffs’ latest relevance theory (coming on the day this brief was due) is that the acquisition value is a proxy for software value because Defendants “never had permission” and therefore “accomplished software piracy on an unprecedented scale.” That may be an argument for copyright remedies, but *Plaintiffs did not bring a copyright infringement claim*.

Plaintiffs also newly argue this case would not exist had Microsoft not purchased GitHub in 2018, apparently on the theories that (1) only due to the Acquisition was OpenAI able to use publicly available code on www.github.com to train the models underlying Copilot, and (2) GitHub would not have developed the GitHub Copilot tool if it had not been acquired by Microsoft years before. These are motive arguments, and have nothing to do with the objective facts of assent, construction, breach, injury, and damages. They are by definition irrelevant.

Moreover, these unalleged assertions are patently false and a big so what? GitHub’s terms of service do not prohibit scraping or otherwise making copies of public code repositories, and in fact OpenAI obtained the code it used to train the models by scraping GitHub’s website. OpenAI

February 23, 2024

Page 3

could have scraped the code whether Microsoft acquired GitHub or not. And, Plaintiffs do not allege breach because somebody copied the code from GitHub, they allege breach because it was sold without attribution when GitHub started offering Copilot. *Why* GitHub developed Copilot (whether or not it was related to the Acquisition) is irrelevant *motive*.

Plaintiffs also hypothesize that due diligence documents may contain interpretations of GitHub's terms of service or the OSS licenses at issue. That is pure speculation. Moreover, the OSS licenses at issue were drafted and published by third parties such as the Free Software Foundation and MIT. Plaintiffs give no reason to believe that the Acquisition encompasses documents defining or explaining terms in drafting such documents, *because Microsoft and GitHub did not draft them*. Recognizing this problem, Plaintiffs' proposed solution is that Defendants go look and see if they can find anything relevant based on Plaintiffs' speculation, and then report back on that investigation. This proposed solution lacks a good faith relevance basis on which to launch an investigation into every document surrounding a \$7.5B acquisition.

Plaintiffs have not satisfied a *prima facie* showing for classwide damages discovery. *See Mantolet v. Bolger*, 767 F.2d 1416, 1424 (9th Cir. 1985), *as amended* (Aug. 27, 1985) (holding that "the plaintiff bears the burden of advancing a *prima facie* showing that the class action requirements of Fed.R.Civ.P. 23 are satisfied or that discovery is likely to produce substantiation of the class allegations" and that "a trial court's refusal to allow class discovery is not an abuse of discretion" absent said "*prima facie* showing"); *Kaminske v. JP Morgan Chase Bank N.A.*, No. SACV 09-00918 JVS, 2010 WL 5782995, at *2 (C.D. Cal. May 21, 2010). The Court has already cast substantial doubt on whether a damages class can ever be certified in this case, holding that only three of the five named plaintiffs have made allegations sufficient to establish standing to bring *any* damages claim. ECF 195. The cases Plaintiffs cite on this point are inapposite, as the discovery Plaintiffs seek here is not discovery needed to establish the existence of a class (i.e., relevant to numerosity, commonality, and typicality), but rather is only related to calculating class damages. There is plenty of time to pursue such discovery if a class is ever certified in this case. The Court has not set a date for summary judgment, let alone pretrial and trial. *See In re REMEC, Inc. Sec. Litig.*, No. CIV 04CV1948 JLS AJB, 2008 WL 2282647 (S.D. Cal. May 30, 2008) (granting defendant's motions for protective order as to third party discovery because requests were outside the scope of the claims).

Plaintiffs argue that Defendants bear the burden of showing that the requested discovery is unduly burdensome and not proportional to the needs of the case, but it is Plaintiffs who must first establish relevance before the warranted degree of burden can be determined. As the Acquisition is irrelevant, the burden of any search is unwarranted.

Defendants' Final Proposed Compromise: Classwide damages discovery be sequenced and conducted only after a class is certified, and the parties be required to negotiate a proportionate production at such time.

III. Plaintiffs' Position

This case would likely not exist at all had Microsoft not bought GitHub in 2018. When it did, Microsoft intended to, and did, seek to extract value from the millions of open-source code repositories hosted on GitHub in ways GitHub had never done as a private company. Especially by leveraging the "exclusive computing partnership" between Microsoft and OpenAI that began in 2019. The first major result was GitHub Copilot, a generative AI product trained by OpenAI,

February 23, 2024

Page 4

marketed by GitHub, and hosted by Microsoft. Most notably, Copilot is derived largely from the work of Plaintiffs and the Class, who own those millions of open-source code repositories (aka “Licensed Materials”), who put them under OSS licenses, and who trusted GitHub’s agreements to abide by the OSS licenses and to never sell their work.

Against this backdrop, Defendants seek a broad protective order designed to prevent Plaintiffs from inquiring into the most significant event in GitHub’s history—its acquisition by Microsoft. (Defendants’ position has changed substantially since their original letter brief to the Court, including three substantive revisions to their arguments). In every iteration of their complaint, Plaintiffs have alleged the relevance of the Acquisition. See, e.g., ECF 200, ¶¶ 5, 187 (“In October 2018, Microsoft acquired GitHub for \$7.5 billion. It was important to Microsoft that programmers use GitHub. Microsoft had developed a well-deserved poor reputation because of its documented ... business practices, including those targeted at open-source programs and programming, and open-source licensing specifically.”).

Nevertheless, Defendants now seek to shield highly probative due diligence and other related evidence possessed by Microsoft and certain third parties. Defendants do so based on a host of misplaced relevancy arguments and legal arguments, raised for the first time here, which might only be proper at summary judgment, after discovery. *See* Rule 56(d). They fail to carry their burden. *See Foltz v. State Farm Mut. Auto Ins. Co.*, 331 F.3d 1122, 1130 (9th Cir. 2003).

Defendants’ core argument—that “[t]he Acquisition ... has nothing to do with the terms of contracts agreed between GitHub and its customers, or between Microsoft and OSS licensors”—is wrong. **The Acquisition constitutes the most direct valuation by Microsoft of the Licensed Materials that belong to Plaintiffs and the Class.** This is a core issue relevant Plaintiffs’ breach-of-contract claims. Regardless of the merit—or lack thereof—on dispositive motions, Defendants fail to bear their heavy burden of preventing access to this information in discovery.

Defendants predicate this argument on the statement that, “[a]ll of the code at issue is by definition licensed without charge...[T]he benefit of the bargain was that Plaintiffs would get credit. The injury is the lack of additional benefit to their reputation.” No. The gravamen of Plaintiffs’ claims is that Defendants never had permission under any contractual theory to use the Licensed Materials to create a commercial AI product like GitHub Copilot. The fact that the Licensed Materials remained available to other users “without charge” for certain uses has no bearing on Defendants’ culpability as license violators, does not provide a limit to Plaintiffs’ remedies, and does not foreclose discovery. Indeed, much of Defendants’ letter brief recapitulates their persistent misunderstandings about open-source licensing—the same misunderstandings that led Plaintiffs to allege that “Defendants have accomplished software piracy on an unprecedented scale.” ECF 200, ¶ 202. Defendants clearly *dislike* the discovery sought by Plaintiffs. But that does not make it irrelevant to the case.

So far, Plaintiffs have sought two categories of documents: 1) due diligence and other documents developed in connection with acquisitions and outside financing of GitHub; and 2) documents relating to the 2018 Microsoft acquisition of GitHub. ECF 200, ¶¶ 5, 187. These documents likely discuss and analyze several important topics: the intellectual-property rights and liabilities of the Licensed Materials; the value of the Licensed Materials; and the risks associated with potential violations of the OSS licenses and GitHub’s terms of service. Other core issues include the extent of GitHub’s rights to the Licensed Materials and other data used for training Copilot; how this data was obtained; and details of the handling of copyright-management information.

February 23, 2024

Page 5

Plaintiffs also expect that access or use of the Licensed Materials is likely addressed in such documents—the third parties subpoenaed by Plaintiffs concede this. Indeed, the materials are likely highly probative on these topics since materials prepared in connection with transactions valued in the hundreds of millions of dollars involving sophisticated parties and outside investors must involve careful consideration of these facts. These witnesses had even agreed to produce a narrowed subset of the material Plaintiffs originally requested—until Defendants interceded. (Defendants have been coy about their behind-the-scenes efforts to inhibit responses by these witnesses. Plaintiffs suggest both parties exchange their communications with third parties.)

Plaintiffs also seek due diligence materials showing the basis for GitHub’s \$7.5 billion purchase because GitHub’s chief asset was its user community and the Licensed Materials hosted on GitHub, including Plaintiffs’ and the Class’ open-source code repositories. The valuations are in fact directly relevant to counter Defendants’ faulty assertion that the Licensed Materials hold no value. *See Mahavisno v. Compendia Bioscience, Inc.*, No. 213CV12207SFCMAR, 2015 WL 5698532, at *5-6 (E.D. Mich. Sept. 29, 2015) (due diligence materials discussing valuation of software for which Plaintiff wrote code were relevant to Defendant’s position that acquiring company did not assign value to such software). Plaintiffs’ interest in the discovery easily outweighs Defendants’ vague suggestions of burden. *In re Coordinated Pretrial Proc. in Petroleum Prods. Antitrust Litig.*, 669 F.2d 620, 623 (10th Cir. 1982).

Plaintiffs are not required to make a *prima facie* showing for classwide damages discovery. *See Brum v. MarketSource, Inc.*, No. 2:17-CV-241-JAM-EFB, 2018 WL 3861558, at *3 (E.D. Cal. Aug. 14, 2018) (distinguishing *Mantolete v. Bolger*, 767 F.2d 1416, 1424 (9th Cir. 1985), *as amended* (Aug. 27, 1985)). The relevancy standard for discovery is even broader than admissibility. *Hickman v. Taylor*, 329 U.S. 495, 501 (1947). Discovery must also encompass any matter that reasonably could lead to others matters that could arise. *See Oakes v. Halvorsen Marine Ltd.*, 179 F.R.D. 281, 283 (C.D. Cal. 1998). Defendants fail to carry their burden to show that discovery should not be allowed or to provide requisite detail for their objections. *Id.*

Defendants’ call for a *prima facie* showing of classwide damages is improper also because the Court has not bifurcated or otherwise sequenced discovery. *See Hall v. Marriott Int’l, Inc.*, No. 319CV01715JLSAHG, 2021 WL 1906464, at *6 (S.D. Cal. May 12, 2021) (collecting California District Court cases overruling assertions from Defendants that discovery relevant to class certification was premature where discovery had not been bifurcated).

Defendants attempt to argue that an unjust enrichment remedy is unavailable for breaches of contract. That is incorrect. Plaintiffs may seek damages and restitution as alternative remedies to their breach of contract claims. *See, e.g., Longest v. Green Tree Servicing LLC*, 74 F. Supp. 3d 1289, 1302 (C.D. Cal. 2015). Moreover, knowing intent is relevant to proving restitutionary disgorgement of profits. *See Kansas v. Nebraska*, 574 U.S. 445, 461 (2015) (“deliberate breach of contract” requires “conscious wrongdoing” in the context of disgorgement of profits under section 39 of the Restatement (Third) of Restitution and Unjust Enrichment).

Plaintiffs’ Final Proposed Compromise: Microsoft and GitHub search the 2018 Acquisition due diligence materials for materials relevant to the issues Plaintiffs seek in the first instance. Following production, the parties to meet and confer to discuss whether further discovery would be appropriate or objectionable.

February 23, 2024
Page 6

Respectfully Submitted,

Dated: February 23, 2024

By: /s/ Annette L. Hurst
Annette L. Hurst (State Bar No. 148738)
ORRICK, HERRINGTON & SUTCLIFFE LLP
The Orrick Building
405 Howard Street
San Francisco, CA 94105
Telephone: 415 773 5700
Facsimile: 415-773-5759
Email: ahurst@orrick.com
Counsel for GitHub, Inc. and Microsoft Corporation

Dated: February 23, 2024

By: /s/ Joseph R. Saveri
Joseph R. Saveri (State Bar No. 130064)
JOSEPH SAVERI LAW FIRM, LLP
601 California Street, Suite 1000
San Francisco, California 94108
Telephone: (415) 500-6800
Facsimile: (415) 395-9940
Email: jsaveri@saverilawfirm.com
Counsel for Plaintiffs and the Proposed Class

L.R. 5-1 SIGNATURE ATTESTATION

As the ECF user whose user ID and password are utilized in the filing of this document, I attest under penalty of perjury that concurrence in the filing of the document has been obtained from each of the other signatories.

Dated: February 23, 2024

/s/ Annette L. Hurst
Annette L. Hurst